Statistical Analysis of Data
Problem sheet #4

Please send by the announced due date to Glen Cowan, Physics Dept., Royal Holloway, University of London, Egham, Surrey, TW20 0EX, or e-mail to `g.cowan@rhul.ac.uk`.

Exercise 4.1: Charged particles traversing a gas volume produce ionization, the mean amount of which depends on the type of particle in question. Suppose a test statistic t based on ionization measurements has been constructed such that it follows a Gaussian distribution centered about 0 for electrons and about 2 for pions, with a standard deviation equal to unity for both hypotheses. A test is constructed to select electrons by requiring $t < 1$.

(a) What is the efficiency for selecting electrons, i.e. the probability to accept a particle if it is an electron.

(b) What is the probability that a pion will be accepted as an electron?

(c) Suppose a sample of particles is known to consist of 99% pions and 1% electrons. What is the purity of the electron sample selected by $t < 1$?

For this exercise you will need the cumulative Gaussian distribution, available e.g. from the CERNLIB routine `FREQ` (C301) or from standard tables.

Exercise 4.2: The number of events observed in electron-positron collisions having particular kinematic properties can be treated as a Poisson variable. Suppose that for a certain integrated luminosity (i.e. time of data taking at a given beam intensity), 3.9 events are expected from known processes and 16 are observed. Compute the P-value for the hypothesis that no new process is contributing to the number of events. To sum Poisson probabilities, you can use the relation

$$\sum_{n=0}^{n_{\mathrm{obs}}} P(n;\nu) = F_{\chi^2}(2\nu; n_{\mathrm{dof}}) \,, \tag{1}$$

where $P(n:\nu)$ is the Poisson probability for n given a mean value ν, and F_{χ^2} is the cumulative χ^2 distribution for $n_{\mathrm{dof}} = 2(n_{\mathrm{obs}} + 1)$ degrees of freedom. This can be computed using the routine `PROB` from the CERN Program Library or looked up in standard tables.

Exercise 4.3: In an experiment on radioactivity, Rutherford and Geiger counted the number of alpha decays occurring in fixed time intervals.[1] The data are shown in Table 1. Assuming that the source consists of a large number of radioactive atoms and that the probability for any one of them to emit an alpha particle in a short interval is small, one would expect the number of decays m in a time interval Δt to follow a Poisson distribution. Deviations from this hypothesis would indicate that the decays were not independent. One could imagine, for example, that the emission of an alpha particle might cause neighboring atoms to decay, resulting in a clustering of decays in short time periods.

(a) Using the data in Table 1, find the sample mean

[1] E. Rutherford and H. Geiger, The probability variations in the distribution of α particles, *Philosophical Magazine*, ser. 6, xx (1910) 698–707.

$$\overline{m} = \frac{1}{n_{\rm tot}} \sum_m n_m m \,, \tag{2}$$

and the sample variance,

$$s^2 = \frac{1}{n_{\rm tot} - 1} \sum_m n_m (m - \overline{m})^2 \,, \tag{3}$$

where is n_m the number of occurrences of m decays and $n_{\rm tot} = \sum_m n_m = 2608$ is the total number of time intervals. The sum extends from $m = 0$ up to the maximum number of decays observed in an interval (here $m = 14$). From $\overline{m}$ and s^2, find the *index of dispersion*,

$$t = \frac{s^2}{\overline{m}} \,. \tag{4}$$

Since $\overline{m}$ and s^2 are estimators of the mean and variance of m (cf. SDA Chapter 5), and since these are equal if m is a Poisson variable, one would expect to find t around 1. One can show that for Poisson distributed m and large $n_{\rm tot}$, $t' = (n_{\rm tot} - 1)t$ follows a χ^2 distribution for $n_{\rm tot} - 1$ degrees of freedom. Furthermore, for large $n_{\rm tot}$, t' becomes Gaussian distributed with mean $n_{\rm tot} - 1$ and variance $2(n_{\rm tot} - 1)$.

(b) What is the P-value for the hypothesis that m follows a Poisson distribution? What set of t (or t') values should be chosen as representing equal or less agreement with the Poisson hypothesis than the observed value?

(c) **(optional)** Write a Monte Carlo program to generate a large number of data sets each consisting of $n_{\rm tot} = 2608$ values of m according to a Poisson distribution. (Poisson random numbers can be generated with the FORTRAN routine `RNPSSN` from the CERN library or from the routine `ranpois.cc` from `www.pp.rhul.ac.uk/~cowan/poisson`.) For the mean value of m, take $\overline{m}$ obtained from the data in Table 1. For each data set, determine t (or equivalently t') and enter its value in a histogram. From the histogram and the value of t obtained from Rutherford's data, Determine the P-value for the Poisson hypothesis. Compare the result to that obtained in (a).

Table 1: Data by Rutherford and Geiger on the number of times n_m that m alpha decays were observed in a time interval of $\Delta t = 7.5$ seconds.

m	n_m	m	n_m
0	57	8	45
1	203	9	27
2	383	10	10
3	525	11	4
4	532	12	0
5	408	13	1
6	273	14	1
7	139	> 14	0